DOCUMENT RESUME

ED 281 879 TM 870 284

AUTHOR Masters, James R.
TITLE 1985-86 Thinking Skills Testing in Pennsylvania's Student Assessment Program.
PUB DATE Jun 86
NOTE 24p.; Paper presented at the Annual Assessment Conference of the Education Commission of the States/Colorado Department of Education (Boulder, CO, June 9-12, 1986).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Abstract Reasoning; Academic Achievement; Age Differences; Cognitive Ability; *Cognitive Processes; *Cognitive Tests; *Decision Making Skills; *Educational Assessment; Elementary Secondary Education; Inferences; *Logical Thinking; Multiple Choice Tests; *Problem Solving; State Programs; Test Construction; Testing Programs; Test Items
IDENTIFIERS Higher Order Skills; Pennsylvania Department of Education; *Pennsylvania Educational Quality Assessment

ABSTRACT

In 1985, for the first time, Pennsylvania's student assessment program included measures of a higher order thinking skills goal termed Analytical Thinking. These tests utilize a decision-making model to assess such skills as drawing inferences, identifying appropriate information to gather before making a decision, analogical reasoning, identifying valid arguments against a conclusion, deductive reasoning and drawing overall conclusions. These skills were assessed at grades 5, 8, and 11 in 1985 and at grades 4, 6, 7, 9, and 11 in 1986. In these testings the instruments were found to evidence internal consistency reliabilities ranging from .79 at grade 4 to .86 at grade 11. The results provided evidence of student growth in thinking skills, since higher percentages of older students than younger students were able to answer the same items correctly. These results were consistent across all age levels and skill areas. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED281879

1985-86 THINKING SKILLS TESTING IN PENNSYLVANIA'S STUDENT ASSESSMENT PROGRAM

James R. Masters
Pennsylvania Department of Education

Presented at the ECS/CDE Assessment Conference

Boulder, Colorado
June 1986

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. R. Masters

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

TM 870 284

1985-86 Thinking Skills Testing in
Pennsylvania's Student Assessment Program

James R. Masters
Pennsylvania Department of Education

Abstract

In 1985, for the first time, Pennsylvania's student assessment program included measures of a higher order thinking skills goal termed Analytical Thinking. These tests utilize a decision-making model to assess such skills as drawing inferences, identifying appropriate information to gather before making a decision, analogical reasoning, identifying valid arguments against a conclusion, deductive reasoning and drawing overall conclusions. These skills were assessed at grades 5, 8 and 11 in 1985 and at grades 4, 6 7, 9 and 11 in 1986. In these testings the instruments were found to evidence internal consistency reliabilities ranging from .79 at grade 4 to .86 at grade 11. The results provided evidence of student growth in thinking skills, since higher percentages of older students than younger students were able to answer the same items correctly. These results were consistent across all age levels and skill areas.

Introduction

In the spring of 1985 the Pennsylvania Department of Education measured Analytical Thinking for the first time in its student assessment program. This measurement was the result of the State Board's 1979 adoption of a new set of 12 Goals of Quality Education to replace the original ten. One of the new goals called for schools to "help every student develop analytical thinking skills".

The Department has been operating a student testing program since 1970 called Educational Quality Assessment (EQA). The intent of this testing effort is to compare the quality of the educational programs in existence in schools by assessing their achievement of the goals established by the State Board. In 1985 it therefore became possible to compare the Analytical Thinking scores of the schools of the state by examining their students' ability to answer test items in this area.

The EQA testing program is voluntary, but each year somewhere between 150 and 225 of the state's 500 districts sign up to participate. For most of the years of its existence, EQA has assessed students in grades 5, 8 and 11. The program focuses upon school means rather than individual student scores. It can be thought of as norm-referenced. Multiple regression is employed to predict how well each school should score, given the type of resources available to it. To stay within budget and testing time constraints, only multiple-choice items are used to measure cognitive achievement areas and matrix sampling is employed. EQA's tests measure schools' achievement of global goals, rather than specific objectives, and the assumption is made that it is only possible to measure a sample of those skills which are included under each goal area.

Figure 1 shows the components a committee of educators offered when they were convened to help define what was meant by Analytical Thinking. As can be

Figure 1
Analytical Thinking Components Listed by a Committee of Educators

The following Program Components contribute toward meeting this goal:

1. Information Management Skills
   a. Identify, locate, utilize and analyze resources relevant to an information need.

   b. Organize and classify ideas and information.

   c. Synthesize information.

2. Logical Thinking Skills
   a. Differentiate between and use inductive and deductive methods of thinking.

   b. Differentiate between and use convergent and divergent methods of thinking.

   c. Apply a sequential thinking process in a problematic situation.

   d. Exhibit confidence in his/her analytical thinking skills.

3. Problem Solving Skills
   a. Formulate a question or problem.

   b. Recognize limits and set goals.

   c. Demonstrate an open-minded and imaginative approach to problem solving.

   d. Identify several alternative solutions to a problem.

   e. Interpret based on observations.

   f. Draw conclusions based on interpretations.

4. Decision-Making Skills
   a. Distinguish between fact and opinion.

   b. Verify judgments on the basis of information, experience and other evaluative criteria.

   c. Recognize biases of all types and determine their influence upon thinking and action.

   d. Analyze and evaluate persuasive statements.

   e. Recognize the need to revise his/her thinking on the basis of new information.

seen by scanning this list, their definition was a broad one, including a variety of thinking skills, including creativity. They also listed attitudinal components and knowledge thought to be basic to being an analytical thinker.

To test such a list with any degree of completeness would have required a great number of items, and to include all of the components within the same tests would have made little sense. The approach taken, then, was to state a model to guide item development and to develop test items for only those skill areas within this model which seemed most amenable to multiple-choice testing. No attempt was made, for example, to deal with the attitudinal and knowledge components listed by the committee. These were certainly felt to be important, but could not be included within tests which focused upon students' abilities to think analytically.

Description of the Analytical Thinking Tests

Development of the Analytical Thinking tests was greatly facilitated by the fact that higher-order thinking skills had been measured within the EQA batteries since 1978. One of the original ten goals focused upon students' abilities to adapt to a world of rapid change. To assess this goal, a decision-making model was stated and thinking skills were assessed using items focusing upon decision situations of central characters. These tests were called measures of "Information Usage".

When development of the Analytical Thinking tests was begun, the Information Usage tests were examined closely with a view toward using them as the basis for the Analytical Thinking tests. The decision-making context seemed to be a good one to maintain and many of the items had performed well over the years. Thus, the Analytical Thinking tests were developed according to a similar model to that outlined for the Information Usage instruments. Test construction was facilitated by the pool of already developed thinking

skills items.

The rationale for using a decision-making context for the items is straightforward. The tests are to measure students' abilities to think analytically. A decision-making situation provides a natural one for such thinking to occur. Students must be able to analyze a great deal of information today if they are to make informed decisions. Asking them to answer thinking questions which are embedded within a decision-making situation is asking them to think within situations similar to those in which they must apply these skills.

In the majority of the items of the tests, then, the facts of a decision situation of a central character or group are enumerated. This character or group must make an important decision. Students are asked to think about the situation and answer multiple-choice questions measuring a variety of thinking skills.

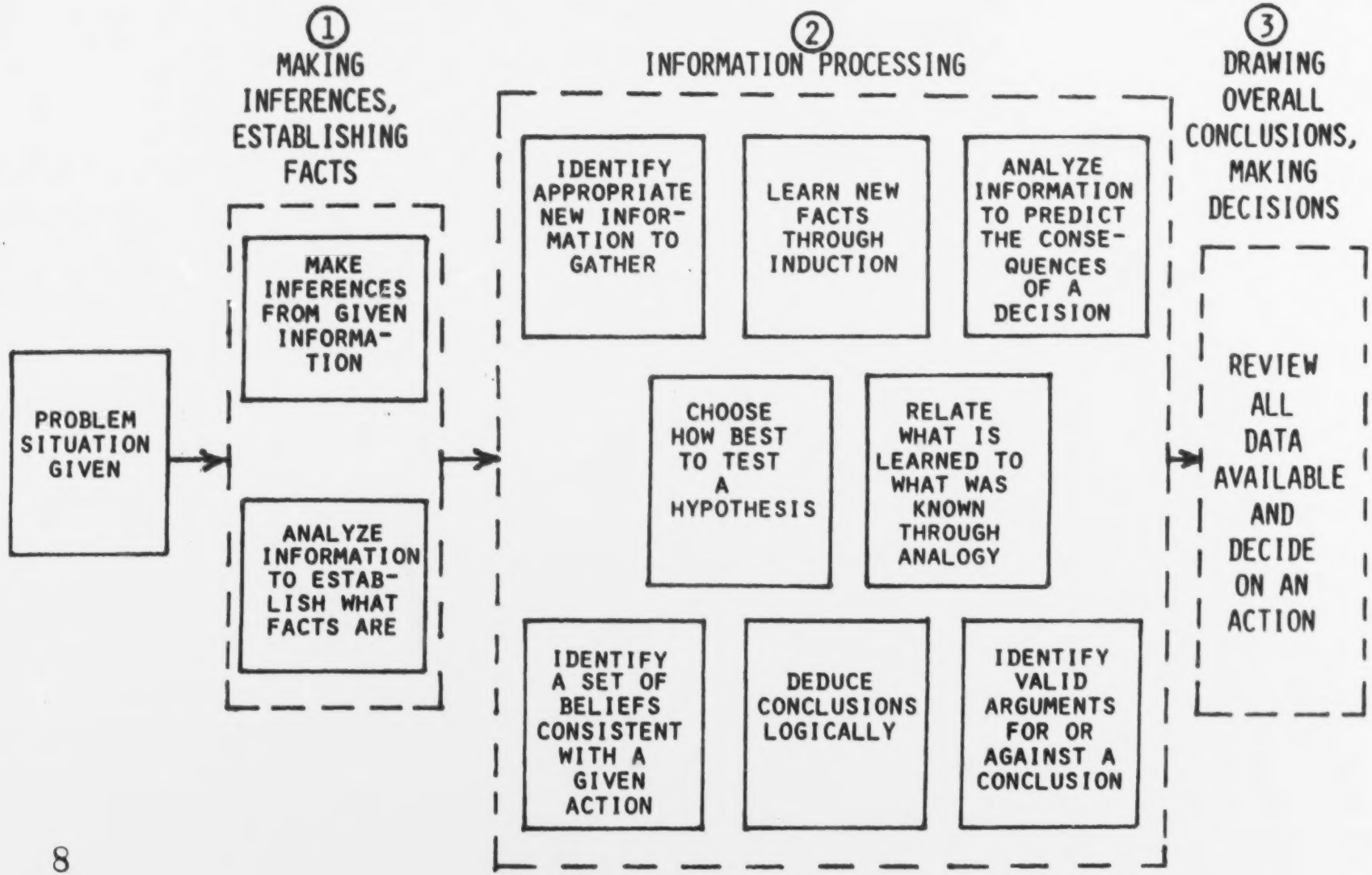
Figure 2 pictures the model which has guided item development. As shown, for most items a problem situation is described. These situations generally focus upon content of interest and relevance to students of the grade level tested. The information presented to students about the situation may be in paragraph form only or, especially at grades 8 and 11, this verbal information may be supplemented by graphs, data tables or information tables.

The first major step shown in Figure 2 is termed Making Inferences, Establishing Facts. The situation has been presented and the student is tested on his or her ability to draw the most probable inferences about what has been given. Also, for very detailed tabular or graphical presentations students are asked to analyze what has been presented as a means of more fully understanding the facts. Thus, the first step shown is one of developing as complete an understanding of the facts and their implications as possible before using them in the decision-making process.

The second major step shown is Information Processing. In this step the facts of the situation are added to or thought about as a means of moving

Figure 2

ANALYTICAL THINKING TESTS: DECISION-MAKING MODEL


1
MAKING INFERENCES, ESTABLISHING FACTS
2
INFORMATION PROCESSING
3
DRAWING OVERALL CONCLUSIONS, MAKING DECISIONS
PROBLEM SITUATION GIVEN
MAKE INFERENCES FROM GIVEN INFORMATION
ANALYZE INFORMATION TO ESTABLISH WHAT FACTS ARE
IDENTIFY APPROPRIATE NEW INFORMATION TO GATHER
LEARN NEW FACTS THROUGH INDUCTION
ANALYZE INFORMATION TO PREDICT THE CONSEQUENCES OF A DECISION
CHOOSE HOW BEST TO TEST A HYPOTHESIS
RELATE WHAT IS LEARNED TO WHAT WAS KNOWN THROUGH ANALOGY
IDENTIFY A SET OF BELIEFS CONSISTENT WITH A GIVEN ACTION
DEDUCE CONCLUSIONS LOGICALLY
IDENTIFY VALID ARGUMENTS FOR OR AGAINST A CONCLUSION
REVIEW ALL DATA AVAILABLE AND DECIDE ON AN ACTION

toward a decision. A variety of skills are measured which have been grouped under this heading. For example, students are asked to evaluate which additional information would be most useful to the central character in reaching a decision, to predict the consequences of various decisions and to choose how best to test a hypothesis.

The final step shown in Figure 2 is Drawing Overall Conclusions, Making Decisions. The major skill tested is the ability to put everything together which is known about the situation and then to choose the most approriate decision to make or action to take.

Two points should be made to clarify more fully what is being attempted through the model being used. First of all, there is no claim being made that the tests measure decision-making ability. Thinking skills are being measured within a decision-making context. Students who receive high scores on these measures would be expected to do well in analyzing their own decision-making situations, but may not be good decision makers because of other factors, such as their propensity to make decisons on the basis of their intuitions.

Second, the model may appear to be picturing decision-making as a rigid, step-by-step process. This is not the intent of the model shown. Obviously when we make decisions we may follow all kinds of paths leading to an overall conclusion.

The 1985 tests based upon this model included 33 items at the grade 5 level and 32 items at each of grades 8 and 11. The following list includes some examples of the types of items contained within the tests:

1) Kim hears an announcer on TV describe a race. She is told that the children who ran fastest ate Crunchies Cereal. Students must help Kim to draw valid inferences about the information presented and to determine which additional information is needed before a conclusion can be drawn. (Grade 5)
2) Mindy dreams that she has found a magic box in the forest. There are a

number of buttons on the outside of the box. She pushes some of these and begins to draw inductive conclusions about the results. She also must relate what she is learning to her past knowledge through analogical thinking. Finally, she must deduce what some other available information means. (Grades 5 and 8)

3) Two major trends are taking place in a town. Students must determine whether enough evidence has been given to warrant a causal relationship between the trends. (Grade 11)

4) A number of central characters are attempting to make decisions about their future careers. A large table of information on five careers is available to them (e.g.,salary level, required education). Given the goals and desires of each central character, students must determine which career is best for each and also what additional information would be most useful to a central character in making his decision. (Grades 8 and 11)

5) Juan has just been given two solid evidences of a fact. However, Jim tells him that this fact can't be true since it doesn't seem that way to him. Students must help Juan deal with this conflict between fact and opinion. (Grade 5)

6) Some club members are trying to choose which product to sell to raise money. They have obtained the past sales records of three products. Students must use the information shown to choose which product(s) would be most logical for the club to sell and also which additional information would be most useful to the club members in making their decision. (Grade 5-small table; grades 8 and 11-graph)

Item Development Methods

Item development for both the Analytical Thinking measures and the

previously developed Information Usage tests involved the creation of a variety of item formats. In the early stages it was determined that attempting to provide the details of a decision-making situation to grade 8 and 11 students in paragraph form was often very difficult and that using a tabular approach was a major improvement. The instruments include information tables, graphs, data tables, figures and paragraphs, all developed to test specific skills contained within the model.

A number of item tryouts were conducted to assess the quality of each item. Determination of which items to retain for the final instruments was a function of a number of types of analyses. Interview sessions were held with students. Also used were classical item analyses, Rasch analyses and critiques by colleagues who examined and responded to the items.

The following list summarizes some of the things we have learned about developing thinking skills items using a multiple-choice format:

1) Interviewing is extremely important in thinking skill item development. Having students respond to items and then describe why they responded the way they did, has added greatly to our ability to locate weaknesses in items and has helped us understand how students of different ages think and respond to test items. Typical item analysis procedures may show that an item is weak but interviewing has the potential for pinpointing what must be changed in the item.
2) If there are no constraints upon accomplishing this, items should be engaging to students. You are asking them to work hard at something and some will just "turn off" if they are not interested. The end result of this is an inaccurate assessment of such students' abilities.
3) There should be a clear rationale for the inclusion of the particular skills in a test. There are many to choose from-- why measure those included? We have found that stating a model and attempting

to create items to measure it has caused us to develop new item types which have worked well for us. Figure 3 is an example of the information table approach we developed to provide the facts of a decision-making situation without having to employ long, complex paragraphs.

4) Because many higher-order thinking skills items require judgments on the part of students, there must be strong assurances that reasonable people would agree about the correct answer to each question. Our tests require students to judge such things as which information would be of the most use to a central character in making a decision. For such items we must be sure that only one answer would be clearly judged the best one. Both student interviews and critiques by colleagues have provided us with such assurances.

5) Item formats must be examined closely to be sure that students will not be choosing their responses because of factors other than thinking. Sometimes aspects of an item can cause students to choose particular responses for reasons unrelated to thinking. For example, they may decide that one of the distractors matches some knowledge they have gained and may choose it on this basis. Other items may be answered incorrectly because students can't read certain words or because they are unable or unwilling to retain the details of a too-complex item long enough to think about them. In an interview session such students may discover on their own that the answer they originally gave was incorrect. The session has caused them to really read the item for the first time.

6) General thinking skills measures should correlate relatively highly with other cognitive measures. That is not to say that correlations of .90 should be viewed as desirable. But, there should not, for example, be a negative connotation to relatively high correlations with reading. Rather, there should be a realization that such

Figure 3

Example of the Information Table Approach

Below is information a school counselor has gathered about five careers. Study this information, then answer the questions below.

Career	% of People Trained For Career Who Find Jobs In That Career	Job Duties	Education Needed	Salary Level	Chance For Advancing To Higher-Level Jobs
I	25%	Work with people almost all the time	4 years of college	Low pay	Excellent
II	50%	Work at desk most of time, with people a small amount	2 years of education beyond high school	Average pay	Fair
III	50%	Work at desk some of time, with people some of time	High school education	Low pay	Fair
IV	Almost 100%	Work at desk some of time, with people some of time	4 years of college	Average pay	Good
V	Almost 100%	Work at desk most of time, with people a small amount	3 years of education beyond high school	High pay	Excellent

1. Kim thinks that, of the five careers listed, Career II is the best one for her. Why might she think this?

   E It would allow her the best chance of getting a job with the least amount of education.
   * F It requires the least education to obtain a job with a salary which is not low.
   G It would allow her the best chance of advancing to a higher-level job with the least amount of education.
   H Career II has the highest pay of the desk jobs.

2. Paul does not want a job which requires him to spend most of his time working at a desk. He wants to be very sure of getting a job in the career for which he has been trained. He would not be satisfied with a low salary. Which career seems best for him?

   A Career I
   B Career III
   * C Career IV
   D Career V

3. Sam has always wanted a job in which he spends most of his time working with people. What is the best question for him to ask about the information in the table?

   E What kind of training would he need for Career II?
   F How soon could he advance if he works in Career V?
   * G If he is trained for Career I, what other jobs could he get?
   H Why do people in Career II receive lower salaries than people in Career V?

correlations will occur routinely when cognitive measures are correlated. A number of years ago we conducted a study in which we correlated our tests with other thinking skills measures, including the Cornell Test of Critical Thinking (Ennis and Millman, 1971). We also correlated these tests with measures of reading comprehension. The correlations among all measures were relatively high. To study these results further, however, we used partial correlation to take out the variance associated with the reading measures from each of the thinking skills correlations. We then examined what was left of the relationships among the thinking measures. We found significant relationships among all of them. Our conclusions were that the measures were assessing something in addition to reading. We assumed that this "something" was thinking. (See Masters and Dorwart, 1980).

Thinking Skill Assessment Results

The Analytical Thinking tests have now been included in the EQA batteries for two assessments, 1985 and 1986. As has been stated, testing took place in 1985 at grades 5, 8 and 11. Because of a grade level conflict with the state competency testing program, assessment was performed in 1986 at grades 4, 6, 7, 9 and 11. Before this change was accomplished grade 4 teachers examined the Analytical Thinking test and all others in the fifth grade battery to make sure they could be used at the fourth grade level without major difficulties. The major changes made in the Analytical Thinking items were simplifications in wording. The teachers believed all of the items would be understandable to fourth graders after these wording changes had been accomplished.

Results from the 1986 testing are still being analyzed, but enough data are available to conclude that the fifth grade tests functioned well at grades 4 and 6, and the eighth grade tests functioned similarly at grades 7 and 9.

Table 1 shows the measurement properties achieved by the Analytical Thinking tests at grades 5 and 8 in 1985 and at grades 4, 6, 7, 9 and 11 in 1986. For all grade levels assessed the tests were estimated to have internal consistency reliabilties of .79 or higher, using a procedure developed by Shoemaker(1973) for tests administered through a matrix sampling approach.

Table 2 was prepared to provide some indications of the validities of the Analytical Thinking tests. In this table the columns labeled "School" contain correlations between school means on the Analytical Thinking tests and school means on the measures shown. The columns labeled "Indiv." show correlations between individual student Analytical Thinking scores and scores of these same students on the other measures. Because there are minor differences in the difficulties of the items on each form, individual student scores were converted to z scores before correlations were computed. Due to the large numbers of students and schools involved in these correlations, all of those shown in this table are significant at the .05 level or at a level less than this.

The areas included in the correlations are of two types. Those six shown in the first group were chosen from the other goal areas tested as examples of how such variables correlated with the Analytical Thinking tests. The remaining variables listed were chosen as relevant ones from among those used to predict goal area scores or to provide additional information to schools about their students or programs.

The magnitudes of the correlations shown do provide some validity support for the Analytical Thinking tests. First of all, the pattern of correlations with the other goal areas is similar to that routinely found for the cognitive areas of the EQA program. Schools which do well in one cognitive area tend to do well in all of them. The individual correlations of the tests with the other goal areas indicate that something is being measured which correlates rather highly (i.e., about .50 to .60) with other cognitive areas, but the magnitude of such correlations is low enough to support the contention that

Table 1
1985-86 Measurement Properties of the Analytical Thinking Tests

Grade	No. of Norm Schools	State Mean	Std. Dev.	Std. Error	Reliab.
4	384	19.33	1.80	0.37	.79
5	361	20.92	1.89	0.36	.82
6	349	22.76	1.89	0.34	.83
7	156	17.29	1.59	0.38	.82
8	138	18.64	1.54	0.38	.83
9	144	19.57	1.79	0.37	.85
11	121	18.28	1.68	0.37	.86

Table 2
1985 Correlations with the Analytical Thinking Tests[1]

Area Measured	Grade 5 School	Grade 5 Indiv.	Grade 8 School	Grade 8 Indiv.	Grade 11 School	Grade 11 Indiv.
Reading Comprehension	.79	.61	.84	.54	.87	.56
Writing Skills	.77	.57	.79	.53	.75	.54
Mathematics	.59	.49	.76	.51	.84	.53
Social Studies	.66	.46	.75	.55	.75	.56
Science and Technology	.67	.53	.76	.45	.86	.47
Self-Concept in School	.33	.26	.33	.26	.38	.26
Interest in School	.20	.15	.29	.19	.44	.21
Percent Low Income Stud.	-.58	NC	-.41	NC	-.41	NC
Parent Educational Level	.51	.17	.43	.19	.43	.19
Parent Expect. for Stud.	NC	NC	.27	.28	.38	.35
Stud.in Academic Program	NC	NC	NC	NC	.36	.38

[1]NC indicates that the area correlated with Analytical Thinking is either not collected at all at the grade level shown or is not collected in a form which could be correlated. For example, Percent Low Income Students is compiled only at the school level and is not capable of being correlated with individual scores.

something different from them is being measured. The fact that correlations of student Self-Concept in School and Interest in School with the tests were lower than were those of the cognitive variables also provides some support for the assumption that the tests measure cognitive skills.

The percentages of a school's population who are from low income families (i.e., Chapter 1 students) and the educational level of a school's parents have always been powerful predictors of the school's achievement in EQA cognitive goal areas. The higher the educational level of the school's parents and the lower the percentage of low income students in the school, the higher the scores typically obtained by the school in such areas. The correlations shown indicate that the Analytical Thinking tests follow similar patterns at the school level. These patterns are ones which have been found repeatedly in this country when cognitive variables have been correlated with socioeconomic variables. The much lower correlations between Analytical Thinking scores and the parent educational level of individual students is also typical of the type of relationship which has routinely been found in Pennsylvania for cognitive measures.

Finally, the last two variables shown are ones for which both school-level and individual-level correlations with Analytical Thinking are of similar magnitude. The degree of parental expectation for student achievement and whether the student is in an academic program in high school would be expected to relate rather highly with performance in cognitive areas, including performance on a thinking skills test.

In addition to examining the measurement properties of the tests, other analyses have been performed as a means of learning as much as possible about the thinking skill performance of Pennsylvania's students and of students in general. One such analysis I will describe involves the study of how student performance differs across grade levels for the same items. This analysis is particularly interesting because the 1985 to 1986 change in the grade levels tested has provided data for seven levels. First, however, I would like to

describe the results we have obtained for one of our eleventh grade items.

The reason I have chosen to discuss the results for this item is that they are rather disturbing, and may point to a problem. We have only one item of this type, so can't make very strong statements, but I would like to learn if others in the country are finding similar results.

The item is rather straightforward. Students are told that in a certain city attacks on older people are decreasing and the number of older people taking self-defense courses is increasing. They are asked to determine what can be concluded on the basis of only this information. The correct answer is that, basically, nothing can be concluded, that two things are happening and neither may be causing the other. Only 25 percent of our eleventh graders answered this item correctly. The remainder attempted to establish causation from the limited information given. For example, 36 percent chose: "Attackers are becoming more afraid of attacking old people".

If this result is not a function of the particular item used and can be replicated widely, one would have to wonder about our students' abilities to determine causality. How often are two things becoming causally related in their minds merely because these appear in the same paragraph or are discussed together?

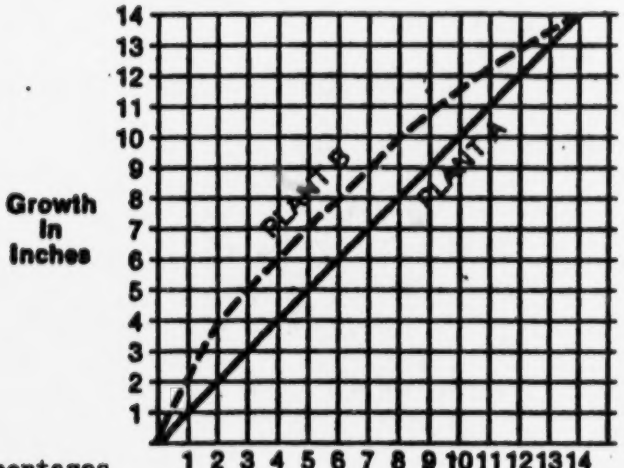
When items used at more than one grade level were studied, the major result found was that greater percentages of older students than younger students were able to answer the same item correctly. For example, for items used at the grade 4, 5 and 6 levels, the average percentages of items answered correctly were 59, 63 and 69, for grades 4, 5 and 6, respectively. Also, for items used at the grade 7, 8 and 9 levels, the average percentages of items answered correctly were 54, 58 and 61, for grades 7, 8 and 9, respectively.

Figure 4 shows some examples of items used across more than three grade levels. The numbers shown are percentages of students who answered the item correctly at each of the grade levels listed. The results to be described are typical of those found for items used at more than one grade level.

Figure 4

Assessment Results for A Sample of the Items
Used at More Than Three Grade Levels

A scientist is attempting to grow larger plants for food. She has graphed the results of her first experiment in Figure 1. Two plants of the same type were grown. Plant A was grown using usual methods. Plant B was grown using a new, experimental method.



Percentages Answering Correctly at each level

1. The scientist is encouraged by the fact that Plant B started out growing faster than Plant A. What would seem to be a good next step for her to take?

G4 58 G5 63 G6 72
G7 75 G8 78 G9 78
G11 81

E Announce her major discovery to the newspapers.
*F Repeat the experiment to make sure the results are the same.
G Do the same experiment again, but this time stop at seven weeks so the experimental plant would be taller.
H Do not do this same type of experiment again, since it was not a success.

Mindy is dreaming. In her dream she finds a large box in the forest. On the outside of it are buttons to push. They look like this:

① Ⓤ Ⓥ ②
③ Ⓦ Ⓧ ④
⑤ Ⓨ Ⓩ ⑥

2. Mindy pushes button Ⓤ and the box plays a tune. She then pushes Ⓥ and the tune becomes slower and slower until it stops. She thinks that pushing Ⓤ is like starting a car and pushing Ⓥ is most like

G4 65 G5 71 G6 77
G7 81 G8 84 G9 87

E blowing the horn.
*F putting on the brake.
G fixing a flat tire.
H starting a lawn mower.

3. A little man says to her: "Only boxes found with less than 7 buttons hold prizes." Mindy sees that the box she has found has 12 buttons. If the little man is right, the box

G4 63 G5 71 G6 75
G7 76 G8 80 G9 80

*E cannot have a prize since it has too many buttons.
F must have a large prize since it has a lot of buttons.
G may have a prize but the little man's message is not of help in knowing this.
H may not have a prize now, but if she takes off 6 buttons it will.

Sid applied for a job as a shoe salesman. He wears his hair long because it is his way of expressing himself. The job interviewer said that he will be hired if he gets his hair cut short. Sid needs the job in order to make his car payments. If he doesn't pay, the bank will take his car. He probably won't be able to find another job in time to save his car.

4. Which of the following are you most sure about?

G7 56 G8 57 G9 59

A None of the salesmen at the store wears his hair long.
B The job interviewer does not like Sid's long hair.
*C Sid does not have the money to make his car payments.
D Sid needs a car in order to travel to work.

G11 59

5. Sid is thinking of some possible ways of solving his problem. For which solution would he be most likely to be able to keep his car and make his car payments?

G7 69 G8 74 G9 75

E Look for a job for which he does not have to shorten his hair.
*F Shorten his hair and keep the job until he makes his payments.
G Ask the people at the bank to let him skip his last payments.
H Shorten his hair only if the store will pay him a higher salary.

G11 77

Vanessa finds a strange-looking box in her attic. It looks something like a music box, but she cannot find an easy way to open it. One side of the box contains buttons which can be pushed. They look like this:

Vanessa begins to push the buttons to try to get the box to play. She pushes button W then button 1 and a song plays. After the song finishes playing she pushes Y then 9 and nothing happens. Here is what happens as she pushes more buttons:

Buttons Pushed	Result
W then 3	Nothing happens
Y then 7	New song plays
X then 4	New song plays

6. What would you expect to happen if she pushes X then 5?

G7 52 G8 57 G9 61

E Same song will play as when W then 1 were pushed.
*F Nothing will happen.
G A new song will play.
H Same song will play as when X then 4 were pushed.

G11 66

7. bcadd is to dcabb as 23144 is to

G7 44 G8 55 G9 61

*E 43122
F 64133
G 46288
H 44132

G11 72

The first item shown, the scientist situation, appears in tabular form at grades 4, 5 and 6, and in graphic form at grades 7, 8, 9 and 11. The wording is also less difficult at the lower levels, but basically the item measures the same thing at all levels. As can be seen, the percentages who answered it correctly ranged from 58 at fourth grade to 81 at eleventh grade.

The next items shown were used at grades 4 through 9. The first item is what I call an "embedded analogy" item, since the analogy is measured within the Mindy situation. Percentages correct ranged from 65 at grade 4 to 87 at grade 9. The second item from the Mindy situation measures deduction. The percentage answering it correctly was lowest at grade 4 (63 percent) and highest at grades 8 and 9 (80 percent).

The fourth and fifth items included in Figure 4 exemplify the short paragraph approach we have used when possible. The items focusing upon Sid's problem were used at grades 7, 8, 9 and 11. The percentages answering each correctly did increase from seventh through ninth grades, but no increase from ninth to eleventh grade was found for item 4, a measure of student ability to draw inferences.

The sixth item shown, the Vanessa item, is one of three developed to measure inductive thinking at the upper levels. The percentages correct ranged from 52 (grade 7) to 66 (grade 11).

The final item shown is one of the few we have included which are not situational. We wished to measure analogical thinking at the upper levels but could think of no effective way to embed such items within a situation. This last Figure 4 item evidenced large differences across grade levels, with only 44 percent of seventh graders able to answer it correctly. At the eleventh grade level, 72 percent could now answer it.

The results we have gathered, then, do demonstrate rather conclusively that some degree of growth in the ability to answer thinking skills items does occur as students grow older. We do not know how much change is taking place for individuals or whether change is limited to only a portion of the student

population. All we have been able to determine at this point is that the percentages who can answer each item correctly are higher at the upper levels than at the lower. We will continue to perform a variety of types of analyses of our data in the future, since Analytical Thinking is now an important component of our EQA testing program.

References

Ennis, R. and Millman, J. Manual for the Cornell Critical Thinking Test, Level X and Cornell Critical Thinking Test, Level Z. Urbana, Illinois: University of Illinois, 1971.

Masters, J.R. and Dorwart, J.P. The Development and Validation of a Decision-Making Skills Test. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Boston, 1980.

Shoemaker, D.M. Principles and Procedures of Multiple Matrix Sampling. Cambridge, MA: Balinger Co., 1973.